https://doi.org/10.5194/egusphere-2026-832
Preprint. Discussion started: 10 April 2026

Atmospheric geopotentials from ERA5 linked to the daily maximum temperature record-breaking in Spain (1960–2023)

Elsa Barrio-Torres[1], Jesús Abaurrea[1], Jesús Asín[1], Jorge Castillo-Mateo[1], Ana Carmen Cebrián[1], and Zeus Gracia-Tabuenca[1]

[1]Department of Statistical Methods, University of Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain

Correspondence: Elsa Barrio-Torres (e.barrio@unizar.es)

Abstract. As the frequency of extreme temperature events increases, so does the need for robust tools to understand them. This work develops and applies a methodological framework to model the occurrence of T_x calendar-day records and their relationship with geopotentials. The analysis includes T_x data from 36 Spanish stations (1960–2023) and geopotentials at 300, 500, and 700 hPa. Exploratory analysis revealed a non-stationary trend in records, a higher frequency in the interior of the peninsula, and decreasing spatial co-occurrence with distance. A hierarchical spatio-temporal logistic regression algorithm prioritizing interpretability was designed. The approach involves: (1) fitting local models per station; (2) applying a spatial consensus filter to reduce initial 1620 parameters to 17 in a base model; and (3) incorporation of interaction terms. Among the tested models, a global model that enhances the base model with geodetic interactions was selected for optimal balance between predictive performance and complexity. Geopotentials at 700 hPa are most relevant for characterizing records, while 300 hPa dominates in the southern corners and 500 hPa in the northern corners. The model demonstrates high predictive accuracy at interior stations, good performance at coastal stations, and adequately reproduces the persistence of record runs and spatial co-occurrence.

1 Introduction

There is an ongoing increase in the rate of extreme climatic events due to global warming. In particular, there is a rising frequency of temperature extremes such as heatwaves and record-breaking events (Saddique et al., 2020; Om et al., 2022; Paredes-Fortuny and Khodayar, 2023; Zhang and Sun, 2025). These events have numerous consequences for ecosystems (Breshears et al., 2021), agriculture (Büntgen et al., 2024), and public health systems (Royé et al., 2021), among others. Europe, specifically, has experienced one of the most pronounced warming trends in recent decades compared to other continents (Hertig et al., 2010; Di Luca et al., 2020; Twardosz et al., 2021), with southern Europe and the Mediterranean sea considered climate *hot-spots*. In these regions, summer temperatures have triggered devastating eventSs such as droughts, wildfires, and "medicanes" (Sousa et al., 2011; González-Alemán et al., 2019; Dupuy et al., 2020).

Although extremes are commonly defined through maxima over a time period or exceedances over a high threshold, record-breaking temperatures, particularly calendar-day records, i.e., the highest value observed on a specific calendar-day at a given station, offer a valuable alternative approach for analyzing extreme behavior. Many meteorological services use these events as climate indices. For example, the U.S. National Oceanic and Atmospheric Administration (NOAA) and its National Centers for Environmental Information (NCEI) provide daily updates on record events across various periods: see Similarly, the Copernicus Climate Change Service (C3S), via the European Centre for Medium-Range Weather Forecasts on behalf of the European Commission, includes these records in its annual European State of the Climate (ESOTC) report; specifically, the 2019 ESOTC report (ESOTC, 2019) analyzed the frequency of calendar-day records and compared it with theoretical expectations under a stationary climate. In Spain, the annual report of the national meteorological service (AEMET) (AEMET, 2023) includes a section (1.1.3) dedicated to the evolution of calendar-day records, distinct from Section 1.1.4, which focuses on heatwaves defined by local thresholds. Moreover, the frequency of record-breaking events is one of the three indicators used to summarize temperature trends in the summary for policymakers of the report.

From a methodological perspective, a major advantage of using record-breaking events is that, in a stationary climate (i.e., no global warming), their occurrence follows well-established, distribution-free probabilistic properties. This makes them a powerful metric for detecting deviations from stationarity and assessing the impacts of global warming (Cebrián et al., 2022).

Recent studies show that the frequency of calendar-day records observed across most of peninsular Spain is inconsistent with a stationary climate. Castillo-Mateo et al. (2025a) found that during the decade 2012–2021, the number of record-breaking temperature events nearly doubled compared to what would be expected under stationary conditions, with the increase being particularly notable during summer. They further revealed that deviations from stationarity are neither temporally nor spatially homogeneous. Their findings include upward trends over time, seasonal variation, persistence effects, and spatial influences such as distance to the coast. Castillo-Mateo et al. (2025b) extended this work to the bivariate analysis of daily maximum and daily minimum temperature records up to 2023, showing different but correlated non-stationarity patterns for records in both series. In China (1960–2023), Cheng et al. (2025) analyzed daily surface air temperature data. They found that summer record-breaking high-temperature events occur more frequently than theoretically predicted, while winter record-breaking low-temperature events exhibit the opposite behavior. They found that this non-stationary behavior has intensified over the last two decades, with post-2020 summers showing a particularly pronounced acceleration in high-temperature record-breaking frequency. None of the previously cited analyses considered potential links to geopotential covariates.

The relationship between extreme temperatures and different atmospheric covariates has been studied in various regions. Capozzi et al. (2025) identified large-scale circulation patterns associated to the occurrence of summer heat waves in the Apennines. Huang et al. (2025) found that abnormal north Atlantic tripole SST anomalies and the Indo-Pacific zonal SST gradient (IPG) were potential origins of the record-breaking extreme temperatures over northern China in October 2023. Qian et al. (2024) investigated the June 2023 record-breaking heatwave in north China and found that insolation and adiabatic heating, linked to anomalous atmospheric circulation (a high-pressure ridge at 500 hPa), played a key role. Ryu and Kang (2023) explored the influence of large-scale variability modes, such as the pacific north America (PNA), north atlantic oscillation (NAO), and pacific decadal oscillation (PDO), on extreme heatwaves in the U.S., using geopotential height and wind variables from the ERA5 database. These teleconnections significantly affected surface air temperatures during their negative phases, due to anomalous anticyclonic circulation patterns. In Spain, García-Valero et al. (2015) examined the impact of sea-level pressure, temperature, and geopotential variables at high atmospheric levels on extreme temperature events. In a similar research line,

https://doi.org/10.5194/egusphere-2026-832
Preprint. Discussion started: 10 April 2026

Castillo-Mateo et al. (2023b) analyzed record-breaking events at different pressure levels over peninsular Spain but found no one-to-one correspondence between surface and upper-atmosphere records. However, to our knowledge, no previous studies have specifically assessed the potential relationship between record-breaking surface temperatures and geopotential covariates.

Understanding and quantifying the influence of geopotential covariates on the occurrence of record-breaking temperatures is essential for conducting climate attribution studies and for developing statistical downscaling methods to project the future frequency of record-breaking temperatures under global warming scenarios. One might argue that such projections could be derived from daily temperature projections, but these often fail to accurately capture the behavior of temperature extremes and, in particular, record events, the most extreme observations. Therefore, generating reliable projections for record-breaking temperatures requires specific downscaling methods (Castillo-Mateo et al., 2025a).

In this context, the objective of this work is to develop a spatio-temporal model that can be used as a statistical downscaling tool for the occurrence of record-breaking daily temperatures. Specifically, we aim to develop a logistic regression model capable of characterizing the occurrence of calendar-day records in daily T_x, using geopotential covariates from the ERA5 reanalysis over a gridded study area. We will provide a new variable selection procedure and tools to fit the model, assess its performance, and draw conclusions from it. We illustrate the procedure by analyzing a dataset of 36 daily temperature series across Peninsular Spain for the summer seasons from 1960 to 2023.

The structure of this paper is as follows: Section 2 describes the dataset, including the temperature and the geopotential series, and Section 3 the exploratory data analysis. Section 4 details the methodology, including the spatio-temporal models and the covariate selection procedure. Section 5 summarizes the findings from the selected model, and Section 6 provides conclusions and outlines future research directions.

2 Study area and dataset

Peninsular Spain is a geographically diverse region. Its complex orography is defined by a central plateau, with the northern part averaging around 1000 meters above sea level and the southern part around 500 meters. The Duero and Tajo river valleys are situated within this plateau, the Guadalquivir valley lies in the south, and the Ebro valley in the north-east. The peninsula also features five main mountain ranges, with peaks exceeding 2000 meters. The Pyrenees form the northeastern border with France, the Cantabrian range lie in the northwest near the Atlantic ocean, the Iberian system occupies the northeast with a northwest-southeast orientation, the Central system divides the Central plateau, and the Baetic range is located in the southeast.

Despite its relatively small area, spanning latitudes 36 to 43.8° N (a north-south distance of approximately 1000 km) and longitudes 9.3° W to 3.° E, peninsular Spain exhibits significant climatic variability over short distances due to its complex topography. The main driver of T_x variation is seasonality, particularly the changes in potential daily solar radiation. During winter, atmospheric circulation over the Atlantic induces a succession of cyclonic and anticyclonic systems over the central and western parts of the peninsula. In contrast, summer is characterized by frequent anticyclonic conditions, leading to reduced cloud coverage and increased solar radiation.

https://doi.org/10.5194/egusphere-2026-832
Preprint. Discussion started: 10 April 2026

The temperature dataset, downloaded from the European Climate Assessment & Dataset (ECA&D) (Klok and Klein Tank, 2009), contains 36 daily T_x series across peninsular Spain for the period 1960–2023, see panel (a) of Figure 1. The selection criteria for these series included a percentage of missing values lower than 0.5%, and ensuring a good representation of the climatic heterogeneity in the study area (Castillo-Mateo et al., 2023b). In this work, we will focus on the calendar-day series derived from the original daily T_x series during the summer months, from June to August (JJA). This results in 92 time series (one per day) across years for each location. Notably, the dataset includes four centennial observing stations recognized by the World Meteorological Organization (WMO): Barcelona-Fabra, Madrid-Retiro, Daroca, and Tortosa [1].

Given the 92 series, one for each day $\ell \in \{1,\ldots,92\}$ of daily T_x across years t (where $\ell = 1$ corresponds to the 1st of June and $\ell = 92$ to the 31st of August), at each location s_i for $i \in \{1,\ldots,36\}$, denoted by $Y_{t\ell}(s_i)$ for $t \in \{1,\ldots,62\}$ (where $t = 1$ is 1960 and $t = 64$ is 2023), the occurrence of calendar-day records is identified by the binary indicators:

$$I_{t\ell}(s_i) = \begin{cases} 1 & \text{if } Y_{t\ell}(s_i) > \max\left\{Y_{1\ell}(s_i),\ldots,Y_{t-1,\ell}(s_i)\right\}, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For $t = 1$, all the observations are defined as a record.

These series are divided into training and testing periods. The training period spans 1960 to 2011 (approximately 80% of the data), while the testing period covers 2012 to 2024. All the models are fitted using the training period, and the testing period is used to validate their performance.

The ERA5 series are downloaded from the *ERA5 hourly data on pressure levels from 1940 to present* collection (Hersbach et al., 2023) at the Copernicus Climate Data Store [2]. We consider daily geopotential series at time 12:00 (denoted as G), measured in m^2/s^2, at pressure levels of 700 hPa, 500 hPa, and 300 hPa for the period 1960–2023. The data were extracted at grid points within the spatial domain [45°N, 10°W, 35°N, 5°E], covering peninsular Spain, using a resolution of $1°\times1°$, resulting in a total of 11×16 grid points, see panel (a) of Figure 1. We denote the geopotential series at each grid point as $Gp.lat.lon$, where p refers to the pressure level and lat and lon represent the latitude and longitude coordinates of the grid point; for example, $G700.41\text{N}.1\text{W}$ refers to the geopotential at 700 hPa at the grid point located at 41°N, 1°W.

3 Exploratory data analysis

The main goal of this exploratory data analysis is to characterize the occurrence of record-breaking temperatures in terms of their temporal evolution, persistence, and spatial dependence. Additionally, we aim to investigate the predictive power of three geopotential levels for the occurrence of daily T_x records. Specifically, we will examine similarities in the underlying trends, patterns in record occurrences, and the distribution of geopotentials conditioned on whether a temperature record has occurred.

---

[1] https://wmo.int/activities/centennial-observing-stations

[2] https://cds.climate.copernicus.eu

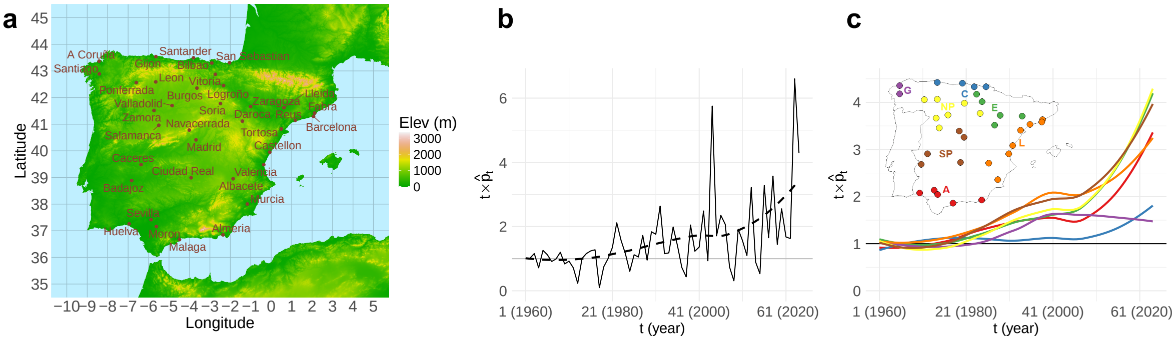


Figure 1. (a): Map of the Iberian Peninsula with the 36 weather stations and the grid were geopotential variables were extracted. (b): Evolution of $t \times \hat{p}_t$ over time, with LOESS smoother (dashed line). Value 1 represents the expected value under the assumption of stationarity. (c): LOESS curves of $t \times \hat{p}_t$ by geographical area: Andalusia (A), Cantabrian coast (C), Ebro valley (E), Galicia (G), Levante (L), North Plateau (NP), and South Plateau (SP). An embedded map is shown with the locations colored by geographical area.

3.1 Exploring record-breaking events in daily temperature

3.1.1 Non-stationary behavior

One of the advantages of using record events is that their occurrence in a stationary series (in this case, a sequence of continuous independent and identically distributed random variables) follows a known, distribution-free probabilistic behavior. Specifically, in a stationary series, the probability of a record at time t is given by $p_t = 1/t$. To assess whether this term adequately captures the temporal evolution of the probability of record, or to identify deviations from stationary behavior, panel (b) in Figure 1 plots $t \times \hat{p}_t$ against t. Note that under stationarity, the expected value of this product is 1. The empirical estimate $\hat{p}_t$ is computed by averaging the indicator variables of records across space and days within each year:

$$\hat{p}_t = \frac{1}{36}\frac{1}{92}\sum_{i=1}^{36}\sum_{\ell=1}^{92} I_{t\ell}(s_i). \tag{2}$$

A LOESS smoother is included in the plot to highlight the underlying trend and deviations from stationarity. A clear non-stationary pattern emerges, with an increasing trend deviating from 1 starting around 1980.

To identify potential spatial differences in the trend, the previous approach is applied to specific geographical areas. Panel (c) in Figure 1 shows the selected regions and the corresponding LOESS curves of $t \times \hat{p}_t$. The results indicate that the northern coastal regions (Galicia and Cantabrian coast) exhibit the smallest increase, whereas inland areas have shown the largest in recent years.

This graphical analysis can be complemented using specific tests to study deviations from stationarity in the occurrence of records (Cebrián et al., 2022; Castillo-Mateo et al., 2023a). Castillo-Mateo et al. (2023b) found that, during the summer months, the tests revealed significant non-stationary behavior in the occurrence of records in the upper tail at most of the 36 stations analyzed in this study, with the exception of those located along the Galician and Cantabrian coasts.

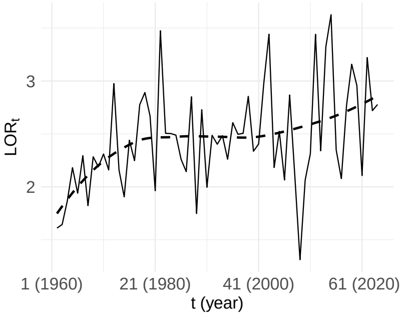


Figure 2. Log-odds ratio, LOR_t, comparing the probability of record given that the previous day was a record, to the probability given that it was not. LOESS smooth in dashed line.

3.1.2 Persistence of record-breaking events

Given the strong serial dependence of daily temperature, some persistence in the occurrence of records is to be expected, that 140 is, the probability of a record on a given day may depend on whether the previous day was also a record. To explore this, we estimate the log-odds ratio which compares the probability of record given that the previous day was a record, $p^{11}_{t\ell} = P(I_{t\ell} = 1 \mid I_{t,\ell-1} = 1)$ to the probability given that it was not, $p^{10}_{t\ell} = P(I_{t\ell} = 1 \mid I_{t,\ell-1} = 0)$ using the expression:

$$\log\left(\frac{p^{11}_{t\ell}/(1-p^{11}_{t\ell})}{p^{10}_{t\ell}/(1-p^{10}_{t\ell})}\right). \tag{3}$$

The empirical log-odds ratio is computed for each year t and shown in Figure 2. Values close to 0 indicate independence, while 145 positive values suggest persistence. The estimates are clearly different from zero, increasing from nearly two at the beginning of the period to almost three by the end, indicating a strong and growing temporal persistence in record occurrences over the years.

3.1.3 Co-occurrence of record-breaking events

The co-occurrence of record-breaking temperatures occurs when two or more locations experience a T_x record on the same 150 day ℓ and year t. To quantify the co-occurrence, we consider, for each day ℓ, the proportion of the 36 observatories having a record on the same day is calculated as

$$PR_{t\ell} = \frac{1}{36}\sum_{i=1}^{36} I_{t\ell}(s_i). \tag{4}$$

Panel (a) in Figure 3 shows these proportions for each day in JJA during the period 1984–2023. The early years, during which the probability of record is relatively high, are omitted to highlight the increasing frequency of record-breaking events. The 155 highest values, approaching 0.8, are observed in recent years, even though the probability of a record in a single series is expected to decrease over time; this suggests that co-occurrence is increasing over time.

https://doi.org/10.5194/egusphere-2026-832
Preprint. Discussion started: 10 April 2026

To explore this possibility, we study the probability that no record occurs at a given day ℓ within year t, in the $N_s = 36$ stations, $P\left(\sum_{i=1}^{N_s} I_{t\ell}(s_i) = 0\right)$. Assuming a stationary climate and independent stations, the probability is calculated as

$$P\left(\sum_{i=1}^{N_s} I_{t\ell}(s_i) = 0\right) = \left(1 - \frac{1}{t}\right)^{N_s}. \tag{5}$$

for any day. Of course, the 36 available time series are not independent even in a stationary climate due to spatial proximity, but this theoretical probability serves as a useful benchmark to evaluate changes in the frequency of simultaneous record events over time. Panel (b) in Figure 3 shows the empirical estimators of these probabilities, calculated as the proportion of days ℓ in a year for which $\sum_{i=1}^{N_s} I_{t\ell}(s_i) = 0$, along with the benchmark under a stationary climate and independent stations. Note that while the reference probability approaches 1 as $t \to \infty$, the estimated probability appears to level off, suggesting an increase in the co-occurrence of record events.

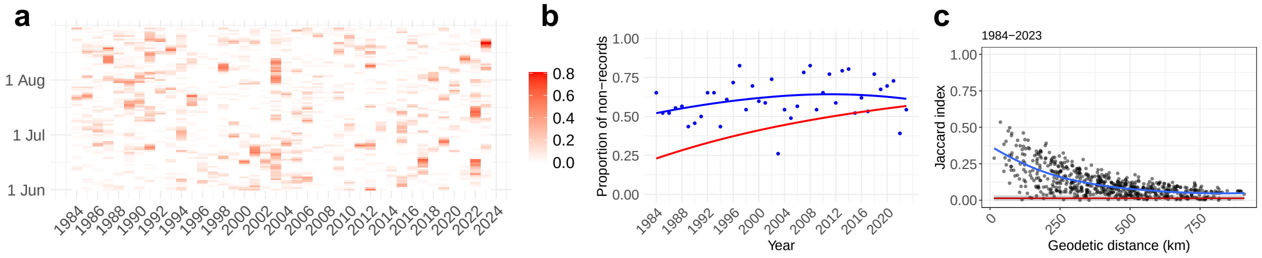


Figure 3. (a): Proportion of days with a temperature record for each day in JJA (rows) across the years in the 1984–2023 period (columns). (b): Proportion of days in each year with $\sum_{i=1}^{S} I_{t\ell}(s_i) = 0$ and LOESS smoother (blue). Expected value under stationarity and independent stations (red). (c): Jaccard index between record series of station pairs for the period 1984–2023 against the distance between stations. A LOESS smoother (in blue) illustrates the overall trend. The expected value and 95% confidence intervals of the Jaccard index for two independent series under a stationary climate (in red and gray) are shown as a benchmark.

Jaccard index. Another useful way to quantify co-occurrence between two series of records is by using the Jaccard index, a similarity coefficient for binary variables. The Jaccard index between two binary series over time indexes t and ℓ, $I_{t\ell}(s_i)$ and $I_{t\ell}(s_k)$, is calculated as

$$J(s_i, s_k) = \frac{\sum_{t,\ell} I_{t\ell}(s_i) I_{t\ell}(s_k)}{\sum_{t,\ell}\left(I_{t\ell}(s_i) + I_{t\ell}(s_k) - I_{t\ell}(s_i) I_{t\ell}(s_k)\right)}. \tag{6}$$

It measures the similarity between the two binary series by comparing the number of observations during which both stations simultaneously register an event (a record, in this case) to the number of observations during which at least one of the stations registers an event. This index quantifies similarity while avoiding the influence of zero-zero coincidences, which do not represent meaningful similarity. The value ranges from 0 (no simultaneous events) to 1 (identical event patterns).

Panel (c) in Figure 3 displays the Jaccard indices for series of records from all pairs of stations during the period 1984–2023, plotted against the distance (in km) between the stations. The expected value and 95% confidence intervals for the Jaccard index between two independent series under a stationary climate were estimated by simulation and are shown in the previous graph, with a mean estimate of 0.01. The plot reveals a strong spatial dependence that diminishes with distance: values are higher than 0.25 on average for distances less than around 100 km, and decrease as the distance increases. For the most distant locations, the binary series can generally be assumed to be independent. However, even among the farthest station pairs, most of the estimated Jaccard indices fall outside the confidence interval obtained under the assumption of records from independent and identically distributed series, providing evidence of deviations from a stationary climate.

3.2 Relationship between geopotential variables and daily temperature

3.2.1 Comparing time trend

To compare the time evolution of the T_x series and the three geopotential variables across the study region, we examine the average of the 36 T_x series and the average of the geopotential variable series at the 11×16 grid points. These averages are computed from standardized signals, where standardization is based on the mean and standard deviation of each series during the 1981–2010 reference period. Figure S.1 displays the LOESS trends of the resulting averages (figures labeled "S" refer to the supplementary material). A similar pattern, marked by an upward trend beginning just before the 1980s, is observed across all four variables, with T_x displaying the steepest slope. Among the geopotential variables, G300 shows the most similar evolution to T_x, and the trend appears to weaken with increasing height.

3.2.2 Comparing evolution in the occurrence of records

In this section, we compare the evolution of the cumulative number of calendar-day records up to year t in T_x and the three geopotential variables. As an illustration, Figure 4 compares the series in the four corners of the grid covering peninsular Spain and the closest stations to those points, A Coruña, Barcelona-Fabra, Murcia and Huelva, together with the expected behavior in a stationary situation, $\sum_{i=1}^{t} 1/i$.

Although a similar temporal evolution is observed across the four variables in each plot, the patterns are not equivalent, with Huelva, and specially Barcelona showing the highest differences. Additionally, there is no homogeneous spatial behavior among the patterns of the four variables. For example, while the highest deviation from stationarity in Barcelona is observed in T_x, this variable shows the lowest deviation in Huelva and Murcia. The behavior is similarly inconsistent in the geopotential variables: in Huelva, the greatest deviation occurs in G300, whereas in A Coruña and Barcelona, G300 exhibits one of the lowest deviations. In summary, although there is a general relationship in the temporal evolution of record occurrences among the four signals, it is not a one-to-one correspondence.

3.2.3 Influence of geopotential variables in the occurrence of record-breaking temperatures

To study the power of geopotential variables to predict the occurrence of record-breaking temperatures, we compare the distribution of the variable given that a record has occurred in T_x or not. Figure S.3 shows the boxplots (without outliers) of the geopotentials at the 300, 500, and 700 hPa pressure levels during the period 1984–2023, conditioned on the occurrence or non-

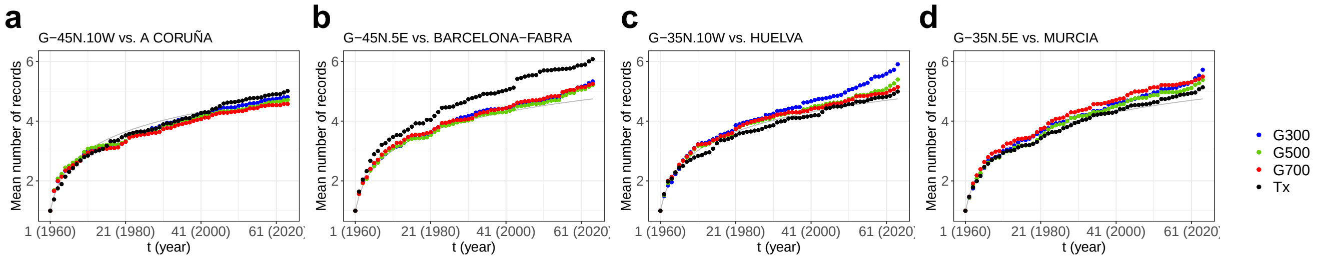


Figure 4. Mean number of records per year t at each corner: (a) NW, (b) NE, (c) SW, and (d) SE of the ERA5 grid for the geopotential in 700 (red), 500 (green), and 300 (blue) hPa. In black the closest station series of T_x records.

occurrence of T_x records. The distribution of geopotential given the occurrence of a record is clearly shifted upward, and its variability is smaller compared to when no record occurs. The difference between the medians for the two conditioning cases is slightly larger in the variables at lower atmospheric levels. In Figure S.4 , the corresponding boxplots for the previously defined geographical zones are also presented. Similar conclusions are drawn, with no significant differences observed between the regions.

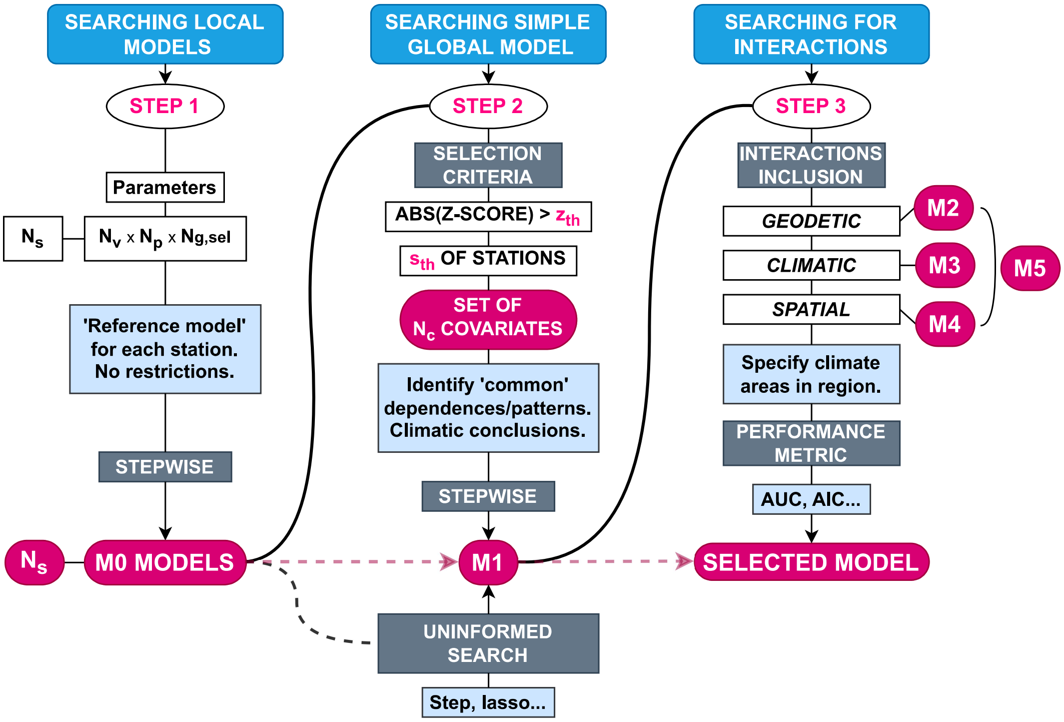
4 Methodology

The aim of this work is to develop a spatio-temporal logistic regression model for the occurrence of calendar-day records in daily T_x, using geopotential covariates from the ERA5 reanalysis over a gridded study area. Model fitting and inference is based on the maximum likelihood framework.

All the statistical analyses were carried out using the software `R`, version 4.4.1 (R Core Team, 2024).

The number of variables that can influence the occurrence of record-breaking temperatures is potentially large. In particular, we consider geopotential variables (G) from $N_v = 3$ pressure levels, evaluated at $N_{g,\text{sel}} = 5$ grid points (the nearest grid point to the station and the four outer corners), and for $N_p = 3$ parameters, including their 1-day lagged terms and second-order polynomial terms. This results in $N_v \times N_{g,\text{sel}} \times N_p$ variables per model (see Figure 5). Exploratory analyses justified this approach, revealing near-perfect correlations among adjacent grid points (Figure S.2), which justified the use of a single representative point. These analyses also identified significant nonlinear trends and persistence effects (Figures 1 and 2), motivating the inclusion of lagged and polynomial terms. This yields 45 potential predictors.

The data set was divided into training ($\approx 80\%$) and testing subsets. The training set comprised the first 51 years of data, while the testing set contained the most recent 13 years. All model computation and feature selection procedures were performed on the training data. The resulting models were then applied to the testing subset and the comparison metrics were computed based on these predictions.



Symbol	Description	Items / Examples	#	Symbol	Description	Items / Examples	#
N_s	Number of stations	See Table S.1	36	$N_{g,sel}$	Grid points used	Corners + nearest	5
N_v	Covariates per grid point	G300, G500, G700	3	z_{th}	z-score threshold	–	2
N_g	Total grid points	11 × 16 grid	176	s_{th}	Station threshold	One-third	12
N_p	Parameters per variable	Linear, lagged, quadratic	3	N_c	Selected covariates	See Figure 6	15

Figure 5. Diagram summarizing the three-step modeling approach, and table describing the notation used.

4.1 A three-step modeling algorithm

A novel three-step procedure is proposed to select appropriate covariates and fit a spatio-temporal model. In the first step, local models are fitted for each station by freely selecting the best combination of covariates from their corresponding set of geopotential-related variables. In the second step, the most prevalent covariates across these local models are identified, and a further covariate selection from this set is performed to obtain a spatio-temporal model based solely on geopotential information. Finally, in the third step, the inclusion of geographical and climatic covariates into the spatio-temporal model is evaluated. These steps are summarized in Figure 5.

Step 1. A logistic regression model is fitted for each of the N_s stations, using the series of daily record indicators $I_{t\ell}(s_i)$ over t and ℓ at each station s_i, as the response variable and $N_v \times N_p \times N_{g,sel}$ variables as potential predictors. Four model configurations were tested and compared: G-terms only; G-terms plus lagged-1 terms; G-terms plus second-order polynomials;

and G-terms plus both lagged and polynomial terms. Covariate selection is performed using a backward stepwise regression based on the standard AIC, that is with $k = 2$ as the multiplier of the number of degrees of freedom. The configuration with the lowest AIC was selected as the best local model. This step yields N_s logistic regression models, one for each station.

Step 2. This step aims to fit a spatio-temporal model for the record indicator series from all stations. Thus, the response variable is obtained by concatenating the response variables from the N_s individual station models. The set of potential covariates for this global model consists of the $N_v \times N_p \times N_{g,sel}$ geopotential-related variables whose absolute standardized coefficients (z-scores) in the local models exceed the threshold $z_{th} = 2$ (approximately equivalent to a two-sided p-value < 0.05) in at least s_{th} stations, where s_{th} is defined as the minimum number threshold (in our case the third part of the N_s local models, $s_{th} = 12$). This guarantees cross-station representativeness. This process resulted in N_c predictors that were used as potential covariates, and a global model was fitted using a stepwise approach starting from the empty model and using $k = 10.83$. This stringent value of k, corresponding to the 99.9th percentile of the χ^2 distribution with 1 degree of freedom, is used to avoid overspecification. The resulting model is denoted M1.

A less time-consuming alternative to Step 2 is to use a non-informative selection approach; that is, to apply the stepwise procedure directly to the original $N_v \times N_p \times N_{g,sel}$ potential covariates.

Step 3. In this final step, more flexible models are considered by extending the set of potential covariates and applying the stepwise approach of Step 2. Four enhanced models are proposed: M2 introduces interactions between geopotential variables and geodetic covariates (latitude and longitude); M3 incorporates interactions with climatic covariates (mean and standard deviation of T_x in a reference period 1981–2010); M4 includes interactions with spatial covariates (altitude and coastal distance); M5 combines all interaction types (geodetic, climatic, and spatial). These models explore different environmental dimensions, enabling the identification of specific geophysical effects while limiting the risk of overfitting.

4.2 Metrics for model comparison

The response variable is binary and exhibits class imbalance (with non-records being predominant). Therefore, model comparison must rely on metrics appropriate for this type of imbalanced classification problems, such as the AUC. The AUC measures the area under the Receiver Operating Characteristic (ROC) curve and it ranges from 0.5 (random prediction) to 1 (perfect classification). Note that all AUC calculations are done using the validation period.

As secondary comparison criteria, we consider: (1) model parsimony (a low number of predictors), (2) Akaike Information Criterion (AIC) values, and (3) AUC performance specifically for coastal stations (those located within 50 km of the sea). This multi-faceted evaluation ensures an appropriate balance between predictive performance and model simplicity while addressing potential geographic biases.

4.3 Model tools

An important step in the modeling process is to evaluate the ability of the selected model to capture key features of record occurrence in a spatio-temporal framework, including spatial co-occurrence and temporal persistence, which is a notable feature of record-breaking daily temperatures in Spain (Castillo-Mateo et al., 2025a).

https://doi.org/10.5194/egusphere-2026-832
Preprint. Discussion started: 10 April 2026

To evaluate the model's ability to reproduce persistence, we compare observed and simulated probabilities of record run lengths, computed as relative frequencies. For each station, we estimate the probabilities of runs of 1, 2, 3, 4, and more than 5 consecutive record-breaking days over the test period. Observed run-length frequencies are obtained by identifying sequences of consecutive records at each station and averaging the resulting relative frequencies across stations. Model-based probabilities are derived from 10,000 simulated binary event series generated using the fitted model. For each simulation, a binary time series is generated by comparing the model's predicted probabilities with a random number uniformly distributed between 0 and 1. A value of 1 is assigned where the prediction exceeds the random number, indicating a simulated record event. From these simulated sequences, the relative frequencies of runs of various lengths are determined. This allows the proposed model to be used as a simulation tool to generate synthetic binary series representing possible trajectories of record occurrences at each location. Then, the simulated and observed frequencies and their corresponding 95% percentile intervals are compared.

To support open science and ensure full reproducibility, all the methods developed in this work are provided as an open-access repository (see https://zenodo.org/records/17055899). This code allows the method to be applied to any region in the world.

5 Results

In this section, we summarize the results of the modeling approach, the local models from Step 1, the global models from Step 2 and Step 3 and a comparison of their performance. Once the best model is selected, some validation analysis is carried out and the type of spatial prediction provided by the model is illustrated.

5.1 Modeling approach

5.1.1 Local models (Step 1)

The stepwise approach applied to each station yields 36 local models including a number of predictors, which ranges from 15 to 35. All models significantly outperform random chance. More precisely, 91.7% of the stations have an AUC greater than 0.8, and 61.1% show an excellent discrimination, with AUC above 0.9; only one, Gijón with AUC$= 0.67$, falls below 0.7. A clear difference in performance appears between inland locations and locations near the coast. The AUC in inland stations is ≥ 0.9 and their median is 0.94. In contrast, stations less than 50,km from the coast have a median AUC of 0.86. This is probably due to the fact that inland environmental conditions are more homogeneous, while coastal stations exhibit higher spatial and temporal variability. Furthermore, they are affected by other factors such as high atmospheric variability, the influence of ocean currents, and the presence of "micro-climates". Details on the fit of the local models are shown in Table S.1.

The table in Figure 6 summarizes, for each predictor, the number of local models in which it is included and the number of models in which its absolute z-score is greater than or equal to 2. Across term types, the frequencies of the 700 hPa predictor are higher than those of the 300 hPa and 500 hPa predictors, suggesting that predictors at higher pressure levels are more informative for predicting the occurrence of daily temperature records. Regarding the grid point, information at the nearest grid point of the 700 hPa geopotential is the most informative: it appears in 26 models with a high z-score in its original form and in 22 models as a lagged effect. Most predictors exhibit a linear effect, as only the quadratic terms of a few variables show a high z-score. These results highlight systematic pressure-level dependencies linked to Iberian topography. As a complement, Figure S.5summarizes the z-scores of the predictors included in the final local models.

	Linear		Lag		Quadratic	
	#	$\#\lvert z\rvert \geq 2$	#	$\#\lvert z\rvert \geq 2$	#	$\#\lvert z\rvert \geq 2$
G300.	17	7	16	10	16	8
G300.45N.10W	9	8	17	9	7	1
G300.35N.10W	28	25	26	20	27	5
G300.45N.5E	7	5	14	7	7	2
G300.35N.5E	33	13	32	22	32	31
G500.	19	6	17	12	18	9
G500.45N.10W	20	15	23	18	18	5
G500.35N.10W	22	14	10	6	21	10
G500.45N.5E	31	23	26	23	29	10
G500.35N.5E	20	14	17	8	19	10
G700.	32	26	25	22	30	10
G700.45N.10W	30	26	19	15	28	16
G700.35N.10W	17	13	9	6	16	3
G700.45N.5E	29	25	22	18	29	15
G700.35N.5E	26	8	12	2	25	20

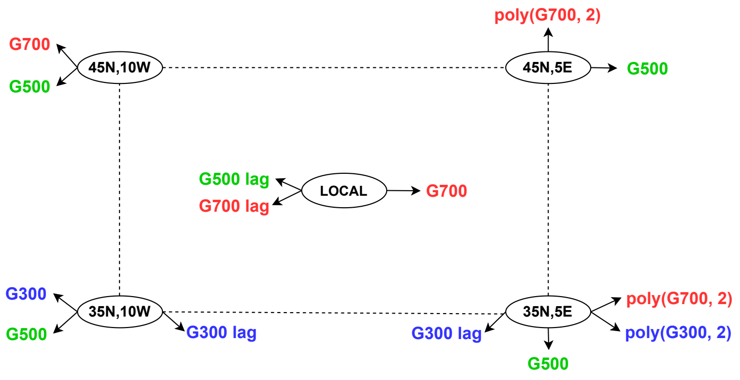


Figure 6. Left: Summary of the predictors included in the local models. # denotes the number of models including each predictor and $\#|z| \geq 2$ the number of models where the corresponding predictor satisfies $|z| \geq 2$. Right: Pseudo-geographical diagram of the geopotential variables in M1 corresponding to the four corners of the grid and the nearest point to a station; “lag” denotes one-day-lagged terms, and “poly” denotes second-order polynomial terms.

5.1.2 Global model (Step 2)

Using a threshold for absolute z-scores greater than or equal to $z_{th} = 2$ in at least one-third of the sites (threshold $s_{th} = 1/3$ out of 36), $N_c = 23$ potential predictors for the global model are selected. Table S.2summarizes the number of predictors obtained under different combinations of thresholds z_{th} and s_{th}. After applying the stepwise procedure to this set of 23 potential predictors, a model including 17 predictors is obtained, see diagram in Figure 6 and S.6. Note that the northern corners include only predictors at 700 and 500 hPa, whereas the southern corners include predictors at 300 and 500 hPa, suggesting a gradient between pressure levels and latitude. Section S.2.2provides more details of the model.

Regarding variable relevance, geopotentials at 700 hPa in the nearest grid point, both the linear and lagged terms ($z = 21.99$, and $z = 20.47$ respectively) show a positive effect on the probability of record, while the NW ($z = -20.67$) and SE ($z = -1.74$ and poly-2: $z = -19.8$) grid points show negative effects. At 500 hPa, the nearest grid point shows a negative effect in the lagged term ($z = -11.71$), whereas the four corner points show positive effects (NW $z = 14.39$; NE $z = 20.68$; SW $z = 24.73$; SE $z = 9.96$). At 300 hPa, we observe a general negative effect in the southern points (SW: $z = -23.53$, lagged term $z = -9.77$, SE: lagged term $z = -12.66$ and a quadratic effect $z = -10.84$, poly-2: $z = 24.15$). Figure S.7summarizes

https://doi.org/10.5194/egusphere-2026-832
Preprint. Discussion started: 10 April 2026

Table 1. Performance metrics for models M1 to M4: number of parameters (k_{coef}), number of covariates (n_v), AIC, the total AUC, inland AUC (AUC$_I$), and for coastal stations total (AUC$_C$), minimum and median AUC, and 10th percentile of the differences between each station's AUC in the global model and in the corresponding local model M0.

Model	k_{coef}	n_v	AIC	AUC	AUC$_I$	AUC$_C$	Min_C	$Q50_C$	$Q10\Delta$
M1	18	15	97808.4	0.86	0.91	0.78	0.59	0.79	0.22
M2	41	21	95606.4	0.88	0.93	0.80	0.64	0.80	0.14
M3	37	20	95995.7	0.88	0.91	0.80	0.67	0.77	0.14
M4	30	18	96856.5	0.87	0.91	0.78	0.60	0.78	0.21

the partial contributions of three relevant predictors (700 hPa, lagged 700 hPa and SW 500 hPa). The results suggest that low geopotentials (cyclonic situations) reduce the probability of record, while high geopotentials (anticyclonic situations) increase those probabilities.

Non-informed stepwise approach. As mentioned above, a simpler alternative to the procedure in Step 2 is to apply a non-informed stepwise approach to the entire set of potential covariates (45 in this case). The model resulting from applying a stopping rule that enforces the same number of predictors as in M1 is summarized in Section S.2.4.The AUC of the non-informed approach decreases to 0.84, compared with 0.86 achieved by M1, and the AUC values at coastal stations are lower (see Figure S.8). This highlights the importance of our informed method for identifying covariates that are consistently relevant along the Iberian coast.

5.1.3 Global model extended with additional covariates (Step 3)

Table 1 summarizes the performance of models M1 to M4 described in Section 4.1. M5 is not included, as its results are identical to M2. The table reports the number of parameters (k_{coef}), number of covariates (n_v), AIC, total AUC, and specific AUC for inland (AUC$_I$) and coastal stations, within 50 km of the coast (AUC$_C$). For coastal stations, where overall performance is lower, additional summary statistics are provided, including the minimum and median AUC, as well as the 10th percentile of the differences relative to the reference local model (M0). Note that the AUC is always calculated using the validation period.

M2 and M3 show the best overall performance, with some subtle differences between them. M3 exhibits a higher minimum AUC, indicating that the lowest-performing coastal station is slightly better in M3. However, M2 achieves better overall global AUC and it is selected as the best model in Step 3. Details of the M2 model are shown in Section S.2.3.

5.1.4 Comparison of models M0, M1, and M2

Comparing M1 from Step 2 and M2 from Step 3, the AUC increases from 0.86 to 0.88, and coastal station values also improve. Figure 7 shows the AUC performance of M2, M1 and M0 (Step 1) for each station, with stations ordered by the AUC values from M0; M0 values (local models) can be considered as a reference value of the best predictive performance attainable with the considered covariates. It is observed that AUC values from M2 are closer to these values than those from M1, specially

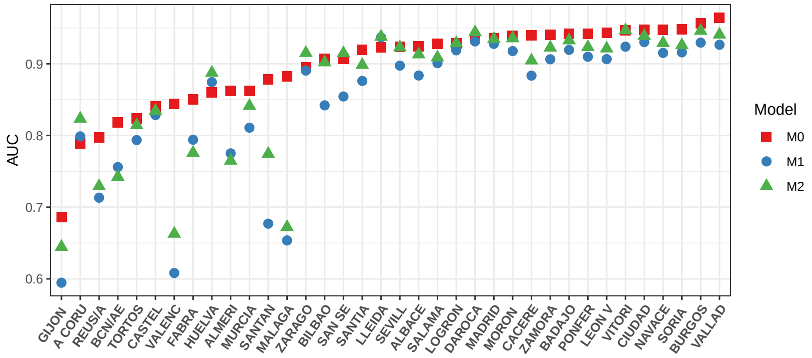


Figure 7. AUCs by station of models M0 (Step 1), M1 (Step 1), and M2 (Step 3), sorted by M0 model's performance, from lowest to highest.

in coastal stations. It is noteworthy, that in some stations, such as A Coruña, Huelva, Zaragoza, San Sebastián, Lleida, and Daroca, M2 outperforms M0. This may occur because the inclusion of additional covariates and/or interactions helps capture systematic patterns that local models fail to exploit, highlighting the potential of appropriately structured global models in certain locations.

The results suggest that including other covariates and interactions is important for effectively modeling coastal stations. Based on these results, M2, the model including as additional covariates, latitude, longitude and necessary interactions is selected as the best model.

5.2 Characterizing climatic properties of heat events

This section summarizes some interesting results and interpretation of the selected model M2.

Figure 8 shows the partial contributions to the model of the G700 covariate at three grid points (45N.10W, 45N.5E, and the nearest grid point to the station) for three example stations (A Coruña, Murcia, and Zaragoza Aeropuerto), which are located in different points of the Peninsula and display distinct climates. The partial contribution of a variable is defined as the value of the linear predictor associated with the terms in which the variable appears, including quadratic terms and interactions. The range of the contribution of G700 at the nearest grid point is generally higher in A Coruña than in Murcia or Zaragoza, indicating that large-scale pressure anomalies at the nearest grid point have a stronger influence on the probability of record occurrence in that point. For G700.45N.10W, the contribution is nearly zero in Murcia, whereas it is strongly negative in A Coruña. This negative contribution suggests that this predictor acts as a moderating term, reducing or counterbalancing the strong positive effect of G700, a behavior already hinted at in the analysis of M1. G700.45N.5E shows negative contributions in both Murcia and Zaragoza, but remains positive, although with a smaller magnitude, in A Coruña. This again reflects a modulatory role: depending on the station's location, this term either lowers or slightly reinforces the influence of G700. A low-pressure situation in the Gulf of Leon might be associated with an influx of warm air from the south (Africa).

The model-based predicted probabilities of record-breaking events are analyzed spatially across a grid covering peninsular Spain, focusing on specific dates across multiple years. To illustrate this, a sequence of maps is produced, each depicting the

https://doi.org/10.5194/egusphere-2026-832
Preprint. Discussion started: 10 April 2026

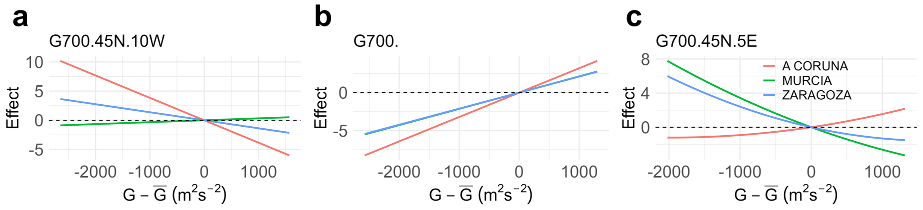


Figure 8. Partial contribution of G700 at (a) grid point NW (45N.10W), (b) the nearest grid and (c) NE (45N.5E) in A Coruña, Murcia, and Zaragoza Aeropuerto stations, model M2. The x-axis are the observed values of each predictor centered around the mean.

predicted probability of a record occurring on different days within a heatwave. These visualizations highlight the model's effectiveness as a tool for spatial prediction of extreme temperature events. As a representative example, panes (a), (b), and (c) of Figure 9 presents the M2 predictions for three days of a heatwave in the validation period (23–25 August 2023). It is observed how the model assigns probabilities exceeding 0.6 in some days and in some areas. These spatial forecasts underscore the model's utility in identifying regions with elevated risk of record-breaking events.

5.2.1 Co-ocurrence of heat events

We check the ability of the statistical model M2 to capture the spatial dependence that leads to simultaneous occurrences at nearby observatories.

First, we analyze the simultaneous occurrence (co-ocurrence) focusing on the 2011–2023 validation period. The Jaccard indices between the binary series representing the occurrence of records for each pair of stations, as described in Section 4.3, are computed over the 2011–2023 validation period. The empirical and model-simulated Jaccard indices across station pairs show a positive correlation of 0.54. Although the simulated values are systematically lower in magnitude, the model is able to capture the main spatial co-occurrence patterns.

Additionally, we evaluated how the selected model predicts extreme heat events that affect a large spatial proportion. For instance, panel (d) in Figure 9 shows the model-based predictions (points) at each station on August 24, 2023, a day on which a record-breaking temperature occurred at most stations, as shown by the contiguous map. Using a true negative rate threshold of 0.95 from the model's ROC curve (correctly classifying 95% of non-record days), displayed on the plot as a reference, most stations where a record occurred that day exceed this prediction threshold. Furthermore, in panel (b) it can be observed that the model's predictions capture a north-south gradient, which agrees with the observed values in panel (e).

5.2.2 Model's ability to reproduce the dynamics of record-breaking events

The model selection was strict, as it relied on the AUC performance obtained during the validation period 2011–2023, which included particularly warm summers. Nevertheless, incorporating atmospheric dynamics as a predictor may help capture features not explicitly represented by other terms.

https://doi.org/10.5194/egusphere-2026-832
Preprint. Discussion started: 10 April 2026

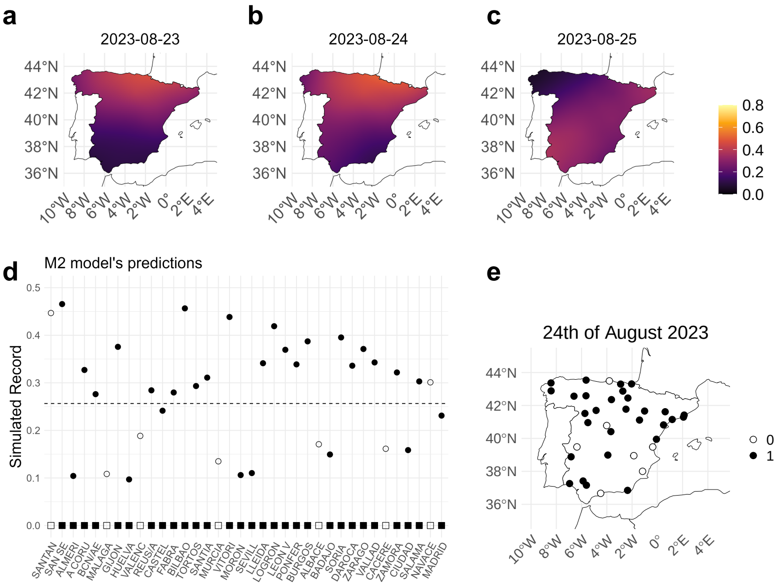


Figure 9. Spatial extreme heat event during August 24th of 2023. (a), (b), and (c): Model-based predicted probabilities of T_x record occurrences. Each panel shows the spatial distribution of probabilities across the grid covering peninsular Spain. The plots correspond to the days (a) 23, (b) 24, and (c) 25 August 2023. (d): scatterplot with the model-based prediction for each station for that day. Dashed line represents the threshold at the true negative rate of 0.95 from the model's ROC curve. (e): observed records (black) and non-records (white) for that day.

In particular, the phenomenon of T_x records is known to exhibit memory or persistence. Castillo-Mateo et al. (2025a) explicitly modeled these characteristics by incorporating an autoregressive component to represent persistence and an additional spatial dependence structure based on spatial Gaussian processes.

Table 2 summarizes the empirical and model-based simulated probabilities of the runs of record-breaking temperatures with different lengths and their corresponding 95% percentile intervals (defined as the 2.5th-97.5th percentiles across stations), over the 2011–2023 validation period. This was previously described in Section 4.3. The results show that the model slightly overestimates the probabilities of single-day runs and underpredicts the occurrence of longer runs ($\geq$5 days). Nevertheless, the 95% percentile intervals for the simulated and observed values are largely comparable, suggesting the model captures the general distribution of runs lengths reasonably well.

Figure 10 shows the model-predicted probabilities of record occurrence during an extreme heat episode in the test period (18–27 August 2023), illustrated for four stations located in different geographical regions. To examine the relationship between record probability and large-scale atmospheric conditions, the scaled values, using the period of 1980–2010, of G300, G500, and G700 at the nearest grid point are also shown. A clear positive influence of geopotentials on record probability is observed, with a slight temporal lag. Furthermore, model M2 successfully reproduces discrete extreme events, such as the separate record runs observed at the Sevilla station.

https://doi.org/10.5194/egusphere-2026-832
Preprint. Discussion started: 10 April 2026

Table 2. Average (95% percentile interval) across stations of the observed and M2-simulated record streaks relative frequencies in the 2011–2023 validation periods.

Runs	Empirical	M2-simulated
Length 1	0.69 (0.53 - 0.90)	0.84 (0.72 - 0.92)
Length 2	0.18 (0.04 - 0.40)	0.12 (0.06 - 0.20)
Length 3	0.06 (0.00 - 0.18)	0.03 (0.00 - 0.07)
Length 4	0.03 (0.00 - 0.11)	0.01 (0.00 - 0.03)
Length 5+	0.03 (0.00 - 0.10)	0.00 (0.00 - 0.02)

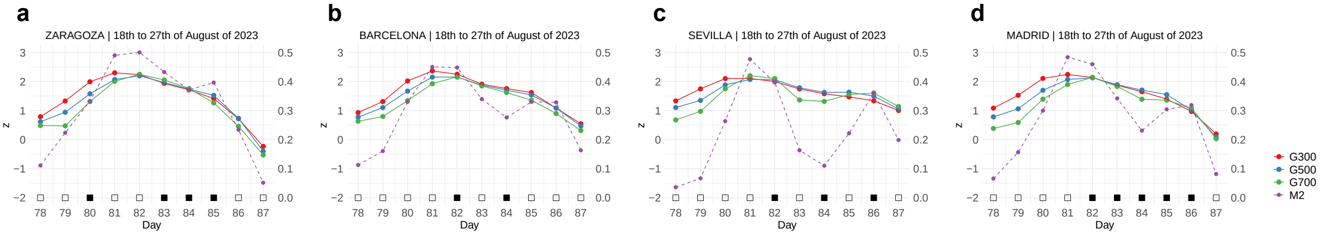


Figure 10. Situation of the atmosphere in the days of the August 18th to the 27th of 2023 in (a) Zaragoza, (b) Barcelona, (c) Sevilla, and (d) Madrid. The boxes at the bottom of the plot are painted in black if the day was a record, and in white if not. Left y-axis represents the geopotentials from the nearest grid point at 700, 500, and 300 hPa scaled using the period 1980–2010 and right y-axis represents the M2 probability of record prediction.

Two main heatwave types were identified. The first corresponds to advection from the east, characterized by a delayed response of T_x records with respect to geopotential anomalies. The second type is associated with advection from the south-west, for which geopotential anomalies and T_x records are synchronized.

The heatwave in Figure 11, occurring at the end of August 2023 corresponds to the advection-from-the-east type. The largest geopotential anomalies are observed on 22 August, with widespread values close to 2. These anomalies are mainly concentrated over the northeastern sector, indicating advection from the European continent. In the following days, geopotential anomalies remain elevated but progressively weaken, and by 25 August, they have largely dissipated. In contrast, temperature observations show that the maximum number of records is reached on 24 August. This temporal mismatch highlights a clear delayed response of temperature records relative to the geopotential anomalies during this event.

A similar behavior is observed during the first and second weeks of August 2003 in Figure S.10. Very high geopotential anomalies are detected on the 4th and 5th of August, originating from the north-east (European continent). From the 6th of August onward, geopotential anomalies begin to dissipate, while temperature records persist for several days, again indicating a delayed surface response.

The heatwave during the second week of August 2021 in Figure S.11 also belongs to this category. Geopotential anomalies are strongest over the southeastern sector, with a pronounced maximum over the western Mediterranean and the Balearic

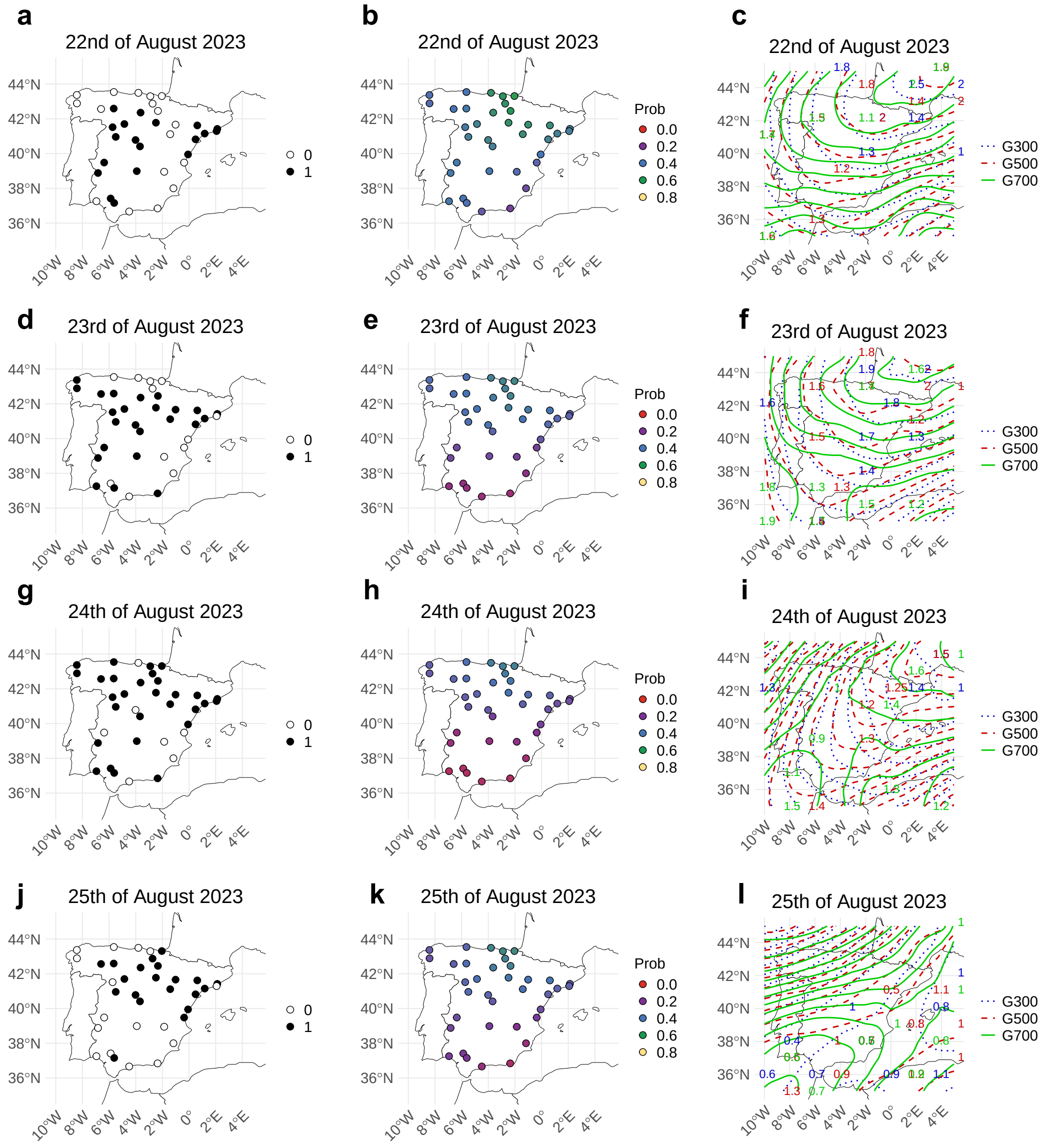


Figure 11. Days of the 22–25 August 2023 (validation period) heatwave. Each row corresponds to a different day. (a), (d), (g), and (j): indicate whether each station experienced a record (black points = records, white points = no record). (b), (e), (h), and (k): show the record probability predicted by model M2 at each station. (c), (f), (i), and (l): display geopotential composites for the specific day at pressure levels of 300 hPa, 500 hPa, and 700 hPa.

Islands. As these anomalies weaken around 14–15 August, the number of temperature records reaches its peak, once more suggesting a delayed effect of surface temperature extremes relative to the large-scale circulation.

In contrast, the heatwave during the second week of August 2023 (Figure S.9) represents the advection-from-the-southwest type. In this case, the largest geopotential anomalies are located over the southwestern sector, and their magnitude increases concurrently with the number of temperature records. Geopotential anomalies and temperature extremes, therefore, evolve together, indicating a more direct coupling between large-scale circulation patterns and surface temperature extremes.

6 Discussion

Castillo-Mateo et al. (2023b) found strong evidence that global warming has increased the number of T_x records in peninsular Spain, particularly during summer. Only the northwestern region remained potentially compatible with stationarity, while 86% of stations exhibited non-stationary behavior. To explore links with atmospheric processes that could support attribution, deviations from stationarity were also analyzed using ERA5 reanalysis data at 850, 700, 500, and 300 hPa. Significant non-stationarity was detected at all levels. However, comparisons between surface stations and nearby ERA5 grid points revealed differences in trends, indicating that surface warming does not always directly mirror warming in the upper atmosphere. Nevertheless, extreme surface temperatures depend on conditions in the overlying atmosphere. This is consistent with our findings, which show different temporal evolutions between geopotential variables and surface records, while confirming that record events are typically associated with higher geopotentials and anticyclonic circulation. The August 2003 heatwave illustrates this behavior: records first appeared in the southwest under moderate anomalies and progressively spread northward as the upper-level anticyclonic pattern intensified, culminating in widespread records in northern regions.

To better characterize the rare occurrence of temperature records, Castillo-Mateo et al. (2025a) introduced a Bayesian hierarchical framework based on binary indicators. This approach goes beyond non-stationarity testing by allowing the estimation of trends, persistence, and spatial patterns. They found that during summer, 98% of Spain exhibits non-stationary behavior, and that global warming has approximately doubled the frequency of record events (90% credible interval: 2.04-2.22). However, their model does not allow direct attribution to changes in atmospheric variability and is relatively complex, involving numerous predictors and random components. Their baseline fixed-effects model achieves an AUC of 0.69 over 1991–2021. In contrast, our simpler baseline model, which includes geopotential covariates, achieved an AUC of 0.86. Castillo-Mateo et al. (2025a) analyzed the August 11–18, 2021 heatwave using a spatio-temporal Bayesian hierarchical model. Their model systematically assigns higher probabilities to record-breaking events. For example, on August 15, probabilities reached values close to 0.9 at several central peninsular Spain stations, whereas our model assigns probabilities closer to 0.6. This discrepancy arises because their approach incorporates spatial random effects, while ours relies exclusively on fixed effects.

Our study represents the first attempt to characterize T_x records across peninsular Spain using geopotential at several pressure levels as predictors. Our results demonstrate a consistent statistical–physical link between upper-level atmospheric circulation and surface temperature records. Exploratory analyses reveal upward trends in geopotentials at 300, 500, and 700 hPa that closely mirror the increase in surface records. Interactions with geodetic covariates modulate these effects: for instance, the influence of the 700 hPa predictor is stronger in western regions and becomes negative in the northwest, acting as a moderating term that counterbalances the positive contribution of the G700 variable at the nearest grid point.

Days when record-breaking temperatures occur, geopotentials are systematically higher at all analyzed levels, indicating the frequent presence of persistent anticyclonic ridges and subsidence that enhance surface warming. These results are consistent with previous studies identifying subtropical ridges and saharian warm-air advection as dominant drivers of Iberian heatwaves (Pfahl, 2014; Serrano-Notivoli et al., 2022; Ventura et al., 2023). In our model, geopotentials at 700 and 500 hPa emerge as the most influential predictors, while 300 hPa geopotentials are particularly relevant in southern regions, highlighting the role of upper-tropospheric dynamics. These relationships reflect anticyclonic subsidence, reduced cloudiness, and enhanced radiative heating at the surface.

A key finding in this work is the marked contrast between coastal and inland stations. Model performance is substantially higher inland (AUC = 0.925) than at coastal locations (AUC = 0.795), confirming the greater difficulty of modeling coastal extremes. Even when additional predictors are added, there appears to be an upper limit to the AUC in several coastal stations, meaning that fully capturing coastal behavior remains challenging and perhaps information about sea surface temperature would be useful. This idea is supported by Pérez-García et al. (2025) who found that most extreme heat events in the Asturian coast coincide with elevated sea surface temperatures. In northwest China, Yang et al. (2026) also found that the interannual variation in the severity of summer cluster high-temperature events is closely related to a simultaneous tripole Sea Surface Temperature anomaly pattern over the north Atlantic.

6.1 Conclusions and future work

Conclusions. This work presents an algorithm for building a model to characterize and predict T_x record events in the peninsular Spain using atmospheric fields (geopotentials) and spatial or climatic covariates. The method employs a local-to-global approach to reduce the dimensionality of the atmospheric predictors, condensing the information of those predictors from 5 grid points into fewer than twenty covariates while maintaining comparable predictive performance and enhancing interpretability. The algorithm further interacts the selected geopotential predictors with spatial and climatic covariates to improve prediction accuracy. This methodology yielded its best performance when geopotential covariates were interacted with geodetic information derived from the latitude and longitude of the target stations, as it was particularly useful in coastal zones of peninsular Spain. The model achieved high predictive accuracy on out-of-sample data (AUC = 0.8787), despite being tested on data from 2011–2023, a period with several very warm summers that made the evaluation especially challenging. Additional validation confirmed the model's ability to reproduce the observed persistence and spatial dependence of record events.

Overall, our parsimonious spatio-temporal logistic regression model provides an interpretable, physically grounded framework that links geopotential variables to surface temperature records. Compared with complex machine-learning approaches, it preserves interpretability while achieving strong predictive performance, highlighting the value of supervised, physically informed variable selection in the study of extreme temperature events.

Future work. The proposed model could be adapted as a statistical downscaling tool to forecast the daily probability of a T_x record and to generate projections of temperature records in a location. This will require to assess the model's downscaling capability by using AR6 model outputs as inputs and producing projections for stations in peninsular Spain over the next 30 years. This application could be explored by using medium-range forecasts of geopotential fields, such as the 24- or 48-hour predictions produced by the ECMWF Integrated Forecasting System, as inputs to the model. This type of tools have already been shown to provide valuable projections of future heatwaves and extreme temperatures in Spanish cities (Abaurrea et al., 2007, 2018).

The method can also be applied to georeferenced binary time series that indicate whether an event occurred at a given location and time, using potential covariates extracted from a grid. This makes the methodology useful in similar contexts where the goal is to examine atmospheric dependencies. For example, Castillo-Mateo et al. (2025b) identified different temporal evolutions in record-breaking for T_x and T_n over peninsular Spain, suggesting that our methodology could be applied to link T_n records to atmospheric variables. Additionally, it would be useful to analyze the relationship between atmospheric variables and the occurrence of extreme daily precipitation events or wind at daily or sub-daily scales. As more grid points are considered around each station, the volume of information will increase both from atmospheric predictors and from topographic variables that influence local precipitation. This added complexity will also raise computational demands, making it necessary to use an optimization algorithm to efficiently build the M3-M5 models, especially when interactions are included.

Code and data availability. The data supporting the findings of this study are openly available from the ECA&D at https://www.ecad.eu/ and ERA5 at https://doi.org/10.24381/cds.bd0915c6. The code to replicate this work is available in the open-access repository https://zenodo.org/records/17055899.

Author contributions. Conceptualization: EBT, ZGT, and J Asín; data curation: EBT and ZGT; formal analysis: EBT, ZGT, J Asín, and J Abaurrea; funding acquisition: AC, J Asín, and ZGT; investigation: EBT, ZGT, J Asín, and J Abaurrea; methodology: EBT, ZGT, J Asín, and J Abaurrea; project administration: J Asín, ZGT, AC, and J Abaurrea; validation: AC and JCM; visualization: EBT and ZGT; writing – original draft: EBT, ZGT, J Asín, and AC; writing – reviewing & editing: JCM and J Abaurrea.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. This work was supported by MCIN/AEI/10.13039/501100011033 and Unión Europea NextGenerationEU under Grant PID2023-150234NB-I00 and RED2024-153680-T Biostatnet; and Gobierno de Aragón under PROY_T21_24-HIDROGIF Grant and Research Group E46_23R: Modelos Estocásticos.

References

Abaurrea, J., Asín, J., Cebrián, A., and Centelles, A.: Modeling and forecasting extreme hot events in the central Ebro valley, a continental-Mediterranean area, Global and Planetary Change, 57, 43–58, https://doi.org/10.1016/j.gloplacha.2006.11.005, 2007.

Abaurrea, J., Asín, J., and Cebrián, A.: Modelling the occurrence of heat waves in maximum and minimum temperatures over Spain and projections for the period 2031–60, Global and Planetary change, 161, 244–260, https://doi.org/10.1016/j.gloplacha.2017.11.015, 2018.

AEMET: Informe anual 2022, http://hdl.handle.net/20.500.11765/15937, 2023.

Breshears, D. D., Fontaine, J. B., Ruthrof, K. X., Field, J. P., Feng, X., Burger, J. R., Law, D. J., Kala, J., and Hardy, G. E. S. J.: Underappreciated plant vulnerabilities to heat waves, New Phytologist, 231, 32–39, https://doi.org/10.1111/nph.17348, 2021.

Büntgen, U., Reinig, F., Verstege, A., Piermattei, A., Kunz, M., Krusic, P., Slavin, P., Štěpánek, P., Torbenson, M., del Castillo, E. M., et al.: Recent summer warming over the western Mediterranean region is unprecedented since medieval times, Global and Planetary Change, 232, 104 336, https://doi.org/10.1016/j.gloplacha.2023.104336, 2024.

Capozzi, V., Di Bernardino, A., and Budillon, G.: Changes in large-scale circulation behind the increase in extreme heat events in the Apennines (Italy), Atmospheric Research, 319, 108 013, https://doi.org/10.1016/j.atmosres.2025.108013, 2025.

Castillo-Mateo, J., Cebrián, A. C., and Asín, J.: RecordTest: An R Package to Analyze Non-Stationarity in the Extremes Based on Record-Breaking Events, Journal of Statistical Software, 106, 1–28, https://doi.org/10.18637/jss.v106.i05, 2023a.

Castillo-Mateo, J., Cebrián, A. C., and Asín, J.: Statistical analysis of extreme and record-breaking daily maximum temperatures in peninsular Spain during 1960–2021, Atmospheric Research, 293, 106 934, https://doi.org/10.1016/j.atmosres.2023.106934, 2023b.

Castillo-Mateo, J., Gelfand, A. E., Gracia-Tabuenca, Z., Asín, J., and Cebrián, A. C.: Spatio-temporal modeling for record-breaking temperature events in Spain, Journal of the American Statistical Association, 120, 645–657, https://doi.org/10.1080/01621459.2024.2427430, 2025a.

Castillo-Mateo, J., Gracia-Tabuenca, Z., Asín, J., Cebrián, A. C., and Gelfand, A. E.: Joint space-time modelling for upper daily maximum and minimum temperature record-breaking, Journal of the Royal Statistical Society Series C: Applied Statistics, p. qlaf046, https://doi.org/10.1093/jrsssc/qlaf046, 2025b.

Cebrián, A. C., Castillo-Mateo, J., and Asín, J.: Record tests to detect non-stationarity in the tails with an application to climate change, Stochastic Environmental Research and Risk Assessment, 36, 313–330, https://doi.org/10.1007/s00477-021-02122-w, 2022.

Cheng, K., Kuang, X., Zhang, Y., and Huang, D.: Quantifying Global Climate Change Impacts on Daily Record-Breaking Temperature Events in China Over the Past Six Decades, International Journal of Climatology e70191., https://doi.org/10.1002/joc.70191, 2025.

Di Luca, A., de Elía, R., Bador, M., and Argüeso, D.: Contribution of mean climate to hot temperature extremes for present and future climates, Weather and Climate Extremes, 28, 100 255, https://doi.org/10.1016/j.wace.2020.100255, 2020.

Dupuy, J.-l., Fargeon, H., Martin-StPaul, N., Pimont, F., Ruffault, J., Guijarro, M., Hernando, C., Madrigal, J., and Fernandes, P.: Climate change impact on future wildfire danger and activity in southern Europe: a review, Annals of Forest Science, 77, 1–24, https://doi.org/10.1007/s13595-020-00933-5, 2020.

ESOTC: European State of the Climate 2019, https://doi.org/10.24381/zw9t-hx58, 2019.

García-Valero, J. A., Montávez, J., Gómez-Navarro, J., and Jiménez-Guerrero, P.: Attributing trends in extremely hot days to changes in atmospheric dynamics, Natural Hazards and Earth System Sciences, 15, 2143–2159, https://doi.org/10.5194/nhess-15-2143-2015, 2015.

González-Alemán, J. J., Pascale, S., Gutierrez-Fernandez, J., Murakami, H., Gaertner, M. A., and Vecchi, G. A.: Potential increase in hazard from Mediterranean hurricane activity with global warming, Geophysical Research Letters, 46, 1754–1764, https://doi.org/10.1029/2018GL081253, 2019.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., et al.: ERA5 hourly data on single levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS)[data set], https://doi.org/10.24381/cds.adbb2d47, 2023.

Hertig, E., Seubert, S., and Jacobeit, J.: Temperature extremes in the Mediterranean area: trends in the past and assessments for the future, Natural Hazards and Earth System Sciences, 10, 2039–2050, https://doi.org/10.5194/nhess-10-2039-2010, 2010.

Huang, H., Zhu, Z., Ma, Q., Hu, S., and Wang, L.: Two oceanic origins for the record-breaking extreme high temperature event over northern China in October 2023, Climate Dynamics, 63, 1–16, https://doi.org/10.1007/s00382-025-07649-0, 2025.

Klok, E. and Klein Tank, A.: Updated and extended European dataset of daily climate observations, International Journal of Climatology, 29, 1182–1191, https://doi.org/10.1002/joc.1779, 2009.

Om, K.-C., Ren, G., Kim, K.-H., Pak, Y.-I., Jong, S.-I., and Kil, H.-N.: Observed trends in extreme temperature events over northern part of the Korean Peninsula during 1960–2019 and a comparative overview, Atmospheric Research, 270, 106 061, https://doi.org/10.1016/j.atmosres.2022.106061, 2022.

Paredes-Fortuny, L. and Khodayar, S.: Understanding the Magnification of Heatwaves over Spain: Relevant changes in the most extreme events, Weather and Climate Extremes, 42, 100 631, https://doi.org/10.1016/j.wace.2023.100631, 2023.

Pérez-García, L., García-Hernández, C., and Ruiz-Fernández, J.: Trends, Atmospheric Patterns, and Spatial Variability of Heatwaves in an Oceanic Climate Area of NW Iberia, Land, 14, 310, https://doi.org/10.3390/land14020310, 2025.

Pfahl, S.: Characterising the relationship between weather extremes in Europe and synoptic circulation features, Natural Hazards and Earth System Sciences, 14, 1461–1475, https://doi.org/10.5194/nhess-14-1461-2014, 2014.

Qian, C., Ye, Y., Jiang, J., Zhong, Y., Zhang, Y., Pinto, I., Huang, C., Li, S., and Wei, K.: Rapid attribution of the record-breaking heatwave event in North China in June 2023 and future risks, Environmental Research Letters, 19, 014 028, https://doi.org/10.1088/1748-9326/ad0dd9, 2024.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2024.

Royé, D., Sera, F., Tobías, A., Lowe, R., Gasparrini, A., Pascal, M., de'Donato, F., Nunes, B., and Teixeira, J. P.: Effects of hot nights on mortality in Southern Europe, Epidemiology, 32, 487–498, https://doi.org/10.1097/EDE.0000000000001359, 2021.

Ryu, J.-H. and Kang, S.-L.: Teleconnections link to summer heat extremes in the south-central US: Insights from CMIP5 and CMIP6 simulations, Weather and Climate Extremes, 42, 100 635, https://doi.org/10.1016/j.wace.2023.100635, 2023.

Saddique, N., Khaliq, A., and Bernhofer, C.: Trends in temperature and precipitation extremes in historical (1961–1990) and projected (2061–2090) periods in a data scarce mountain basin, northern Pakistan, Stochastic Environmental Research and Risk Assessment, 34, 1441–1455, https://doi.org/10.1007/s00477-020-01829-6, 2020.

Serrano-Notivoli, R., Lemus-Canovas, M., Barrao, S., Sarricolea, P., Meseguer-Ruiz, O., and Tejedor, E.: Heat and cold waves in mainland Spain: Origins, characteristics, and trends, Weather and Climate Extremes, 37, 100 471, https://doi.org/10.1016/j.wace.2022.100471, 2022.

Sousa, P. M., Trigo, R. M., Aizpurua, P., Nieto, R., Gimeno, L., and Garcia-Herrera, R.: Trends and extremes of drought indices throughout the 20th century in the Mediterranean, Natural Hazards and Earth System Sciences, 11, 33–51, https://doi.org/10.5194/nhess-11-33-2011, 2011.

Twardosz, R., Walanus, A., and Guzik, I.: Warming in Europe: recent trends in annual and seasonal temperatures, Pure and Applied Geophysics, 178, 4021–4032, https://doi.org/10.1007/s00024-021-02860-6, 2021.

585 Ventura, S., Miró, J. R., Peña, J. C., and Villalba, G.: Analysis of synoptic weather patterns of heatwave events, Climate Dynamics, 61, 4679–4702, https://doi.org/10.1007/s00382-023-06828-1, 2023.

Yang, X., Zhou, B., Zhang, X., Xie, W., Tu, Y., and Cheng, Y.: Interannual relationship between the severity of summer cluster high-temperature events in Northwest China and the North Atlantic sea surface temperatures, Climate Dynamics, 64, 12, https://doi.org/10.1007/s00382-025-07987-z, 2026.

590 Zhang, Y. and Sun, Y.: Attribution of historical extreme heat events in different climate zones of China, Environmental Research Letters, 20, 034 013, https://doi.org/10.1088/1748-9326/adb2a8, 2025.